(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 749 109 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
18.12.1996 Bulletin 1996/51

(51) Int Cl.[6]: **G10L 5/06**

(21) Application number: **96850108.0**

(22) Date of filing: **04.06.1996**

(84) Designated Contracting States:
CH DE DK ES FR GB IT LI NL

(30) Priority: **16.06.1995 SE 9502202**

(71) Applicant: **TELIA AB**
123 86 Farsta (SE)

(72) Inventor: **Lyberg, Bertil**
610 72 Vagnhärad (SE)

(74) Representative: **Karlsson, Berne**
Telia Research AB,
Rudsjöterrassen 2
136 80 Haninge (SE)

(54) Speech recognition for tonal languages

(57) The present invention relates to a method and device at speech-to-text conversion. From a given speech the fundamental tone is extracted. A model of the speech is further created from the speech. In the model a duration reproduction in words and sentences is obtained. The duration reproduction is compared with a segment duration in the speech. From the comparison is obtained information which decides which type of accent that exists. at which a text with sentence accent information is produced.

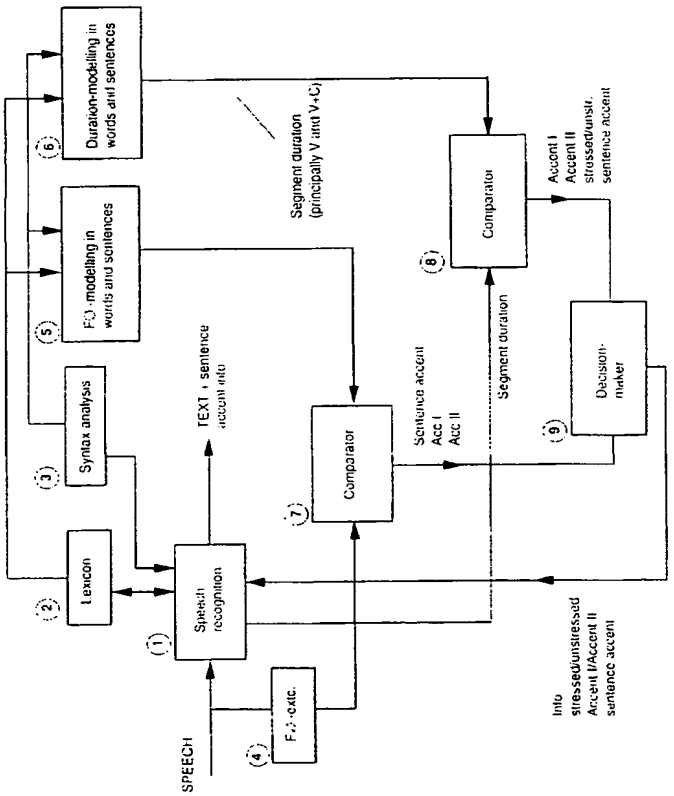


Figure 1

EP 0 749 109 A2

Description

TECHNICAL FIELD

The present invention relates to speech-to-text-conversion. Especially is referred to the possibility to analyse a given speech and from this obtain information about different accents as well as stressed and unstressed syllables in a speech. This information is of importance at interpretation of the given speech.

PRIOR ART

In the speech recognition systems which are utilized at present with for instance HMM. fundamental tone and duration information are regarded as disturbances. Information regarding sentence accent types, respective stressed or unstressed syllables have in the known applications been performed on the basis of statistical methods. The information which is obtained via the accentuation in the speech has at that not been possible to identify.

In patent document US 5220639 is described a speech recognition at mandarine chinese characters. A sequence of single syllables is recognized by separate recognition of syllables and mandarine tones and putting together recognized parts for recognition of the single syllable under utilization of hidden markov models. The recognized single syllable is used by a markov chinese language model in a linguistic decoder section for determination of corresponding chinese character. A tone pitch frequency detector is utilized. The tone pitch frequency detector detects characteristics regarding the pitch frequency of the unknown signal and transmit it to one for the tone recognition included personal computer, in which Markov model probabilities for the five different tones are calculated.

In patent document US 4852170 is described language translation under utilization of speech recognition and syntesis. Each speech segment is logically analysed for identification of phoneme class affiliation. After that the frequency spectrum of the segment is analysed for identification of specific phonemes within the type.

In patent document US 4489433 is described a speech information transmission by means of telex equipment. After the transmission. speech data can be converted into a readable message of characters. The technology according to the document is principally intended to be applied at the Japanese language. The accent type of Japanese words is a tone pith accent and can be identified depending on the position of the point of time between the syllables at which the tone pitch frequency is changed abruptly to a low frequency. The word accent code indicates a sudden change in tone pitch and fundamental tone frequency. usually caused by the accent of a special syllable in a word.

Patent document US 4178472 deals with a speech instruction identification system which suggests commands by examination of a pattern of changes in syllable sounds. The fundamental tone frequency is used as a symbolic value for speech/sound.

Patent document EP 180047 relates to recognition of spoken text and following printing. For each segment of the recognized spoken text a corresponding string of characters is stored. Lexical information is utilized.

DESCRIPTION OF THE INVENTION

TECHNICAL PROBLEM

At speech recognition there is a need to identify different sentence accents and stressed respective unstressed syllables in words and sentences. Methods or devices to generally appoint different types of accent respective stressed/unstressed syllables have so far been lacking. The prosodic information has so far not been used at speech recognition but is regarded as a disturbance at the statistical methods which are used. The prosodic information is necessary at advanced speech understanding systems at speech-to-speech translation. By analysing the prosodic information and appointing the location of the accents and the types of the accents in words and sentences is obtained an increased understanding of the given speech and a possibility to translate it better between different languages. Problem further exist to appoint stressed/unstressed syllables in words and sentences. By the ability to identify the location of stressed respective unstresed syllables in words and sentences is also given an increased possibility to identify the real meaning of a sentence. Consequently there exists a need to identify said parameters and utilize these in connection with speech recognition.

The aim with the present invention is to indicate a method and device for identification of the proper sense of a given speech.

THE SOLUTION

The present invention relates to a method at speech-to-text conversion where the fundamental tone is extracted from a speech. From the speech is created a model of the speech. From the model is obtained a duration reproduction in words and sentences. The duration reproduction is compared with a segment duration in the speech. From the comparison is decided which type of accent that exists and a text with sentence accent information is produced. Sentence accents of type 1 and 2 are discernible. Further. stressed and unstressed syllables are discernible. From the model a model is modelled of the fundamental tone in words and sentences. The invention further indicates that the fundamental tone is compared with the modelled fundamental tone. at which indication for possible accents is obtained. The possible accents at the comparison of the fundamental tone and the comparison of duration are compared. and decision is made which type of accent

or stressed/unstressed syllable that exists. The decision is utilized to adjust the model. A text is at that produced which with great probability obtains one with the speech corresponding meaning. At the creation of the model lexical information is utilized. The lexical information indicates alternative accents in the words. The lexical information further indicates alternative durations for different segments in the words which are recognized. Syntax analysis of the model is utilized at modelling of the fundamental tone in the sentence. The syntax analysis of the model is utilized at the modelling of the sentences.

The invention further relates to a device at speech-to-text conversion. A fundamental tone is extracted from a speech in a speech recognition equipment. A model of the speech is created in the speech recognition equipment. From the model a duration reproduction in words and sentences is created. The device further is arranged to compare the duration reproduction with a segment duration in the speech. Decision regarding type of accent is performed in the device on the basis of the comparison. A text with sentence accent information is produced. Sentence accents of type 1 and 2 are discernible. as well as stressed and unstressed syllables. From the model a model of the fundamental tone in words and sentences is produced. The extracted fundamental tone is compared with the modelled fundamental tone and an indication of possible locations of accents is obtained. The possible accents at the fundamental tone comparison are compared and decision is made regarding which type of accent or stressed/unstressed syllables that exist. The decision is utilized for correction of the model, and a text is produced which with great probability corresponds with the meaning of the speech. Lexical information is utilized at the creation of the model. In the lexical information is included information about different types of accents respective stressed/unstressed syllables etc in different words and sentences. By means of the lexical information alternative accents and accent locations are obtained in the words which have been obtained from the lexical information. Alternative durations for different segments in the recognized words are obtained from the lexical information. At modelling of the fundamental tone in sentences. a syntax analysis of the model is utilized. At modelling of the sentences the syntax analysis of the model is utilized.

ADVANTAGES

The invention allows that a prosodic information is utilized at speech analysis, at which an increased understanding of the speech is obtained. The increased understanding will increase the possibility to utilize spoken information in different connections. for instance translation from a speech into another speech at automatic speech translation. The invention further allows an increased possiblity to utilize spoken information in different connections for control of different services in a telecommunications network. at control of different devices, computers etc.

DESCRIPTION OF FIGURES

Figure 1 shows the invention in the form of a block diagram.

DETAILED EMBODIMENT

In the following the invention is described on the basis of the figures and the terms therein.

A produced speech is fed into a speech recognition equipment. 1. In the speech recognition equipment the speech is analysed in its components. At this different recognized sequences appear, which are made up to words and sentences. The analysis which is performed in the speech recognition equipment is performed with for the professional in the field wellknown technology. Consequently, for instance Hidden Markov Models, HMM. can be utilized. In this type of analysis the fundamental tone and the duration information are regarded as disturbances. Information regarding the duration of the segments is however possible to derive in the Markov model. By the analysis in the speech recognition equipment are obtained a number of recognized sounds which are put together to words and sentences. One consequently obtains a set of combinations of syllables which are possible to combine to different words. Said words consist of words which exist in the language, respective words which do not exist in the language. In a first check of the recognized words, possible combinations are transferred to a lexicon. 2. The lexicon consists of a normal lexicon with pronounciation and stress information. In the lexicon different possible words are checked, which can be created from the recognized speech segments. From the lexicon information. information about the possible words which can exist based on the recognized speech is fed back. In the speech recognition equipment after that a compilation of the words is made to clauses and sentences. This information is transferred to a syntax analysis. 3. In the syntax analysis is checked whether the suggestions to clauses and sentences which have occured are, from a linguistic point of view acceptable or not acceptable in the language. The lexical and syntactical information is after that transferred to a fundamental tone modulating unit 5. and a duration modulating unit. 6. In the fundamental tone modulating unit the fundamental tone is modulated on the basis of the lexical and syntactical information. At that a fundamental tone modulation in words and sentences is obtained. The obtained information is transferred to a comparator. 7. which also obtains an information regarding the fundamental tone of the speech which has been extracted i the fundamental tone extractor. 4. At the comparison in 7 information about possible locations of the sentence accent, accent 1 and accent 2 is obtained.

From the lexical and syntactical analysis a model of the duration in words and sentences is also produced. At that the lexicon contains information about the duration of different syllables in the possible words which have been obtained at the analysis of the speech. In syntax analyses also possible durations for different parts in the sentences which are possible and in the different words are produced. From the total information a segment duration, where the duration of the wowels and possibly following consonants are the most important, is obtained. The in this way obtained information is transferred to a second comparator, 8. The comparator, 8, also obtains an information segment duration in the real speech from the speech recognition equipment. At the comparison in the comparator, 8, information about possible locations for accent 1, accent 2, stressed or unstressed syllables and sentence accents is obtained. This information is transferred to a decision-maker, 9, which has also received information from the first comparator, 7, regarding sentence accent 1 and accent 2 from the fundamental tone information. The decision-maker after that compiles the information from the two comparators and decides whether accent 1, accent 2, stressed or unstressed syllable or sentence accent exists. The obtained information is after that fed back to the speech recognition equipment which modifies the original model and after that feeds out a text with sentence accent information.

By the suggested solution a possibility is given to recognize a speech and reproduce it in a correct way with better accuracy than in previously known methods. The in the original speech given meaning can at that be reproduced in a correct way. Further the information can be utilized in the case the given speech shall be translated into another language. Further possibility is given, in a correct way to find right word and expression and determine which of alternative meanings that shall be utilized at the analysis of words and sentences. The uncertainty at previous methods, principally statistical methods, to decide the proper sense of different words, is by the suggested method reduced in a drastical way.

The invention is not restricted to the in the description presented embodiment, or by the patent claims, but can be subject to modifications within the frame of the idea of invention.

Claims

1. Method at speech-to-text-conversion, at which a fundamental tone is extracted from a speech, and from the speech a model of the speech is created, **characterized** in that from the model a duration reproduction in words and sentences is obtained, that the duration reproduction is compared with a segment duration in the speech, that from the comparison is decided which type of accent that exists, and that a text with sentence accent information is produced.

2. Method according to patent claim 1, **characterized** in that accent 1, accent 2, and sentence accents are discerned.

3. Method according to patent claim 1, **characterized**in that stressed and unstressed syllables are discerned.

4. Method according to any of the previous patent claims, **characterized** in that from the model a model of the fundamental tone in words and sentences is modelled.

5. Method according to any of the previous patent claims, **characterized** in that the extracted fundamental tone is compared with the modelled fundamental tone at which indication for possible accents are obtained.

6. Method according to any of the previous patent claims, **characterized**in that the possible accents at the fundamental tone comparison and the duration comparison are compared and decision is made which type of accent or stressed/ unstressed syllable that exists.

7. Method according to any of the previous patent claims, **characterized** in that the decision is utilized for correction of the model at which the produced text with great probability gets one with the speech corresponding meaning.

8. Method according to any of the previous patent claims, **characterized** in that at the creation of the model lexical information is utilized.

9. Method according to any of the previous patent claims, **characterized** in that the lexical information indicates alternative accents in the words.

10. Method according to any of the previous patent claims, **characterized** in that the lexical information indicates alternative durations for different segments in the words which are recognized.

11. Method according to any of the previous patent claims, **characterized** in that syntax analysis of the model is utilized at modelling of the fundamental tone in the sentence.

12. Method according to any of the previous patent claims, **characterized** in that the syntax analysis of the model is utilized at modelling of the sentences.

13. Device at speech-to-text conversion, and a fundamental tone is extracted from a speech in a speech

recognition equipment. and a model of the speech is created in the speech recognition equipment. **characterized**in that from the model a duration reproduction in words and sentences is created, the device is further arranged to compare the duration reproduction with a segment duration in the speech. decision regarding accent type is performed in the device on the basis of the comparison. and a text with sentence accent information is produced.

14. Device at speech-to-text conversion according to the patent claim 13, **characterized** in that accent 1 and accent 2, and sentence accents are discernible.

15. Device at speech-to-text conversion according to the patent claim 13, **characterized** in that stressed and unstresed syllables are discernible.

16. Device at speech-to-text conversion according to the patent claims 13 up to and including 15, **characterized** in that from the model a model of the fundamental tone in words and sentences is produced.

17. Device at speech-to-text conversion according to the patent claims 13 up to and including 16. **characterized** in that the extracted fundamental tone is compared with the modelled fundamental tone och that an indication for possible locations of accents is obtained.

18. Device at speech-to-text conversion according to the patent claims 13 up to and including 17. **characterized** in that the possible accents at the fundamental tone comparison are compared and that decision is taken regarding which type of accent. or stressed/unstressed syllable that exists.

19. Device at speech-to-text conversion according to the patent claims 13 up to and including 18. **characterized** in that the decision is utilized for correction of the model and a text is produced. which with great probability corresponds to the sense of the speech.

20. Device at speech-to-text conversion according to the patent claims 13 up to and including 19. **characterized** in that lexical information is utilized at the creation of the model.

21. Device at speech-to-text conversion according to the patent claims 13 up to and including 20. **characterized** in that alternative accents in the words are obtained from the lexical information.

22. Device at speech-to-text conversion according to the patent claims 13 up to and including 21. **characterized** in that alternative durations for different segments in the recognized words are obtained from the lexical information.

23. Device at speech-to-text conversion according to the patent claims 13 up to and including 22. **characterized** in that at modelling of the fundamental tone. a syntax analysis of the model is utilized.

24. Device at speech-to-text conversion according to the patent claims 13 up to and including 23. **characterized** in that, at modelling of the sentences. the syntax analysis of the model is utilized.

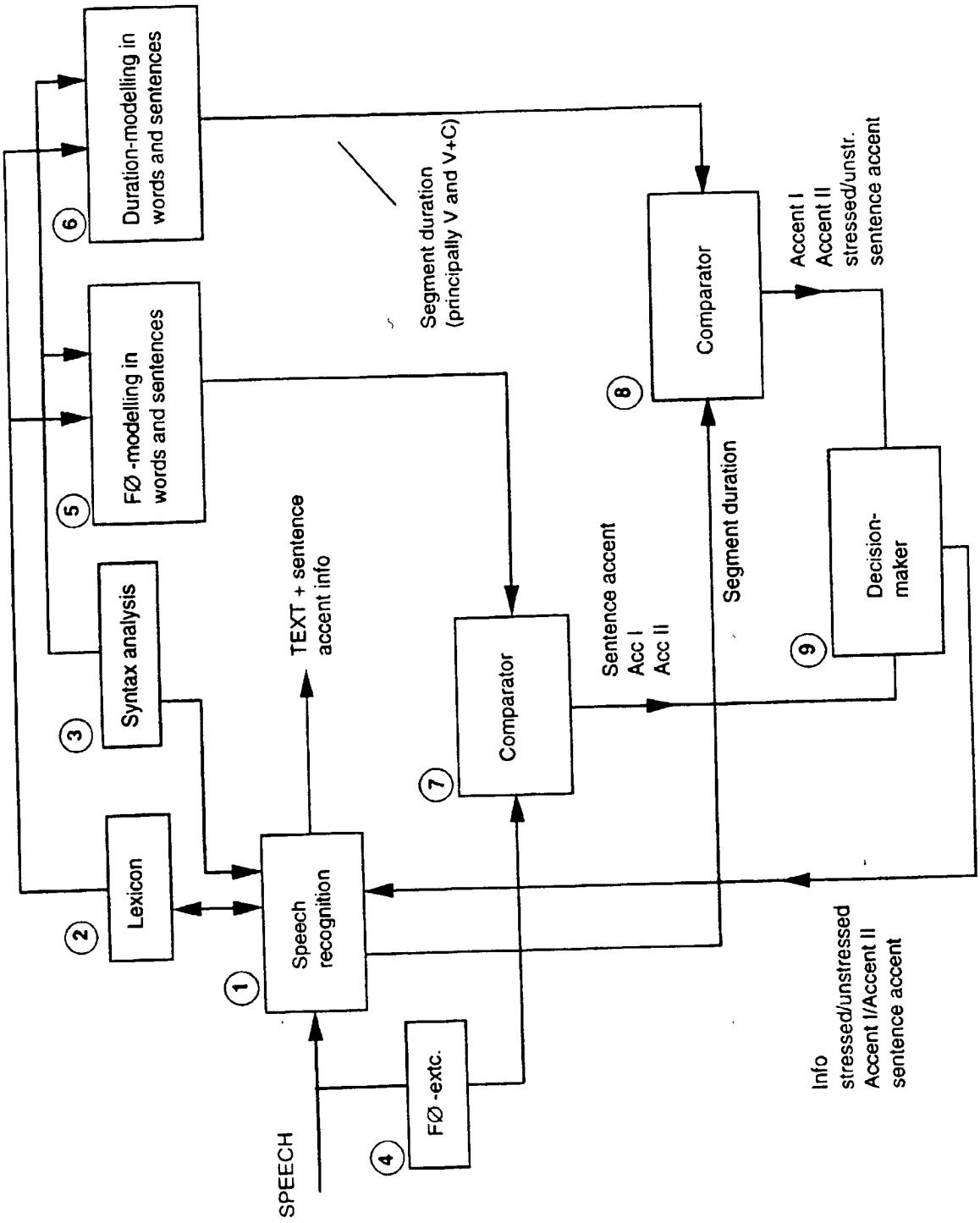

Duration-modelling in words and sentences
6
FØ -modelling in words and sentences
5
Syntax analysis
3
Lexicon
2
Speech recognition
1
FØ -extc.
4
SPEECH
TEXT + sentence accent info
Segment duration (principally V and V+C)
Comparator
7
Sentence accent Acc I Acc II
Comparator
8
Segment duration
Accent I Accent II stressed/unstr. sentence accent
Decision-maker
9
Info stressed/unstressed Accent I/Accent II sentence accent

Figure 1

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 749 109 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
29.04.1998 Bulletin 1998/18

(51) Int Cl.[6]: **G10L 5/06**

(43) Date of publication A2:
18.12.1996 Bulletin 1996/51

(21) Application number: **96850108.0**

(22) Date of filing: **04.06.1996**

(84) Designated Contracting States:
CH DE DK ES FR GB IT LI NL

(30) Priority: **16.06.1995 SE 9502202**

(71) Applicant: **TELIA AB**
123 86 Farsta (SE)

(72) Inventor: **Lyberg, Bertil**
610 72 Vagnhärad (SE)

(74) Representative: **Karlsson, Berne**
Telia Research AB,
Rudsjöterrassen 2
136 80 Haninge (SE)

(54) Speech recognition for tonal languages

(57) The present invention relates to a method and device at speech-to-text conversion. From a given speech the fundamental tone is extracted. A model of the speech is further created from the speech. In the model a duration reproduction in words and sentences is obtained. The duration reproduction is compared with a segment duration in the speech. From the comparison is obtained information which decides which type of accent that exists. at which a text with sentence accent information is produced.

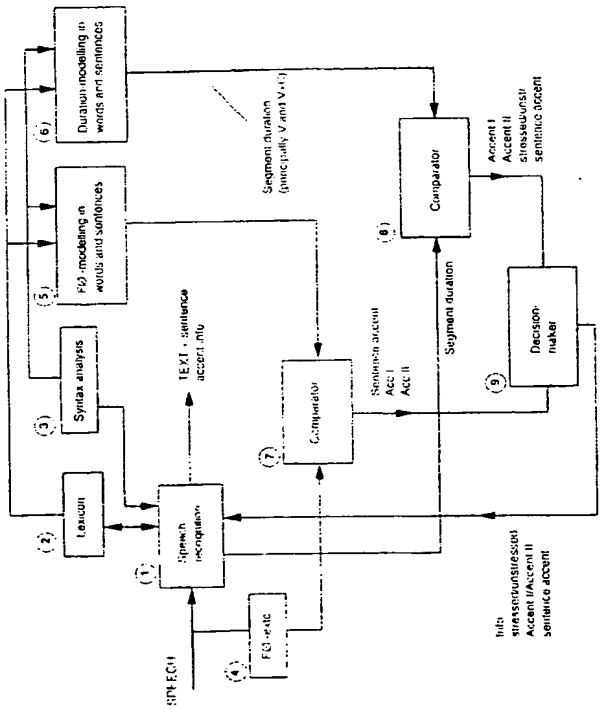


Figure 1

EP 0 749 109 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 85 0108

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int Cl.6)
X	DATABASE WPI Week 9534 Derwent Publications Ltd., London, GB; AN 95-256367 XP002058208 & SE 9 303 623 A (TELIA AB), 4 May 1995 * abstract * -& EP 0 664 537 A (TELIA AB) 26 July 1995 * column 5, line 17 - column 6, line 27; figures 1,2 *	1,4,7,8, 11-13, 16,17, 19,20, 23,24	G10L5/06
P,X	WO 96 00962 A (TELIA AB; LYBERG BERTIL (SE)) * page 7, line 7 - page 8, line 23; figure 1 *	1,2,4-7, 13,14, 16-19	
A	FUJISAKI H ET AL: "ANALYSIS AND MODELING OF WORD ACCENT AND SENTENCE INTONATION IN SWEDISH" SPEECH PROCESSING, MINNEAPOLIS, APR. 27 - 30, 1993. Vol. 2 OF 5, 27 April 1993, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. pages II-211-214, XP000427763 * page II-211, right-hand column, line 1 - page II-214, right-hand column, line 10; figures 1-6 *	1-6, 13-16	

TECHNICAL FIELDS SEARCHED (Int Cl.6)
G10L

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	9 March 1998	Pulluard, R

CATEGORY OF CITED DOCUMENTS

X: particularly relevant if taken alone
Y: particularly relevant if combined with another document of the same category
A: technological background
O: non-written disclosure
P: intermediate document

T: theory or principle underlying the invention
E: earlier patent document, but published on, or after the filing date
D: document cited in the application
L: document cited for other reasons

&: member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 85 0108

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	FUJISAKI H ET AL: "MANIFESTATION OF LINGUISTIC AND PARA-LINGUISTIC INFORMATION IN THE VOICE FUNDAMENTAL FREQUENCY CONTOURS OF SPOKEN JAPANESE" 18 November 1990 , PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP), KOBE, NOV. 18 - 22, 1990, VOL. 1 OF 2, PAGE(S) 485 - 488 , ACOUSTICAL SOCIETY OF JAPAN XP000503413 * page 486, right-hand column, line 6 - page 488, right-hand column, line 22; figures 2-7 *	1,4,13, 16	
A	BRUCE G ET AL: "ANALYSIS AND SYNTHESIS OF DIALOGUE PROSODY" 18 November 1990 , PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP), KOBE, NOV. 18 - 22, 1990, VOL. 1 OF 2, PAGE(S) 489 - 492 , ACOUSTICAL SOCIETY OF JAPAN XP000503414 * page 490, left-hand column, line 11 - right-hand column, line 43 *	1,2,4, 13,14,16	

TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	9 March 1998	Pulliard, R

CATEGORY OF CITED DOCUMENTS

T: theory or principle underlying the invention
E: earlier patent document, but published on, or after the filing date
D: document cited in the application
L: document cited for other reasons
& member of the same patent family, corresponding document

This Page Blank (uspto)